

Universidade de Brasília

Modelo de regressão log-linear com fração de cura

Camila Leal

Brasília
2013

Camila Leal
Bacharel em Estatística

Modelo de regressão log-linear com fração de cura

Orientadora:
Profa. Dra. **Juliana Betini Fachini**

Brasília
2013

DEDICATÓRIA

À minha mãe,

Selma B. do Rêgo Leal, por todo apoio, amor e dedicação. Por sempre acreditar em mim. Por nunca ter medido esforços para que eu concluísse mais esta etapa com todo conforto e êxito.

Ao meu pai,

Antônio Celso Leal, (in memoriam), Por ter sido meu exemplo como profissional e pessoa. Por ter sonhado com este momento. Por tudo que conseguiu me ensinar em 12 anos. Por ainda se fazer presente.

AGRADECIMENTOS

À Deus, por tudo que me proporcionou durante toda a minha vida. Por a cada dia renovar minhas forças para conclusão desta etapa.

À professora, orientadora e amiga Juliana, por toda paciência, carinho e atenção. Por ter me acolhido desde o primeiro e-mail. Por sempre me compreender e por ter auxiliado muito mais que devia e eu esperava.

Aos professores do Departamento de Estatística da Universidade de Brasília que durante os últimos 4 anos fizeram toda a diferença para a minha formação profissional.

À minha mãe Selma pelo amor, carinho, incentivo e paciência neste período.

Ao padrasto lindo Júlio pelo amor e carinho de pai. Por sempre acreditar e depositar sua confiança em mim.

Ao meu irmão Daniel pela preocupação, carinho e paciência.

Às minhas tias Sandra e Sônia por todo amor de mãe que sempre me deram.

À minha mais que amiga Nananda pela paciência, apoio, alegrias e momentos de descontração.

Ao Diego Armando por todo apoio, carinho, paciência, palavras de incentivo e principalmente por todo amor.

À querida amiga Thavana por me acompanhar desde o início, por ter compartilhado das mesmas dores e alegrias e principalmente por todo apoio e amparo durante a execução deste trabalho.

À eterna galerinha mais ou menos Marcus, Snoop, Tomás e Paty pelo convívio diário, amizade, companheirismo, risadas e descontrações nos momentos de desespero.

Aos demais colegas de curso pela convivência e pelo apoio nos momentos de dificuldade.

Às minhas amigas Cinthia, Danielle, Jéssica, Simara, Karol, Thais, Flávia, Luana, Babi, Vanessa e Monique pela amizade, pela diversão e pela paciência com a minha ausência.

Aos meus chefes e ex-chefes Murilo, Paloma, Leonardo, Erick e Igor por todo conhecimento repassado, pela confiança, pela flexibilidade e pela paciência.

A todos que de alguma forma contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO	9
1 INTRODUÇÃO	11
2 REVISÃO DE LITERATURA	13
2.1 Notação e conceitos básicos	13
2.1.1 Tempo de falha	13
2.1.2 Censura	14
2.1.3 Representado o tempo de sobrevivência	16
2.1.4 Estimador de Kaplan Meier	17
2.1.5 Função Total do Teste	18
2.2 Distribuições de Probabilidade para os Dados	19
2.2.1 Distribuição Beta Burr XII (BBXII)	19
2.2.2 Distribuição Burr XII (BXIII)	20
2.2.3 Distribuição Log-logística	20
2.3 Estimação dos Parâmetros pelo Método de Máxima Verossimilhança	21
2.4 Teste da Razão de Verossimilhança	22
2.5 Fração de Cura	23
2.6 Modelo de Regressão Locação e Escala	24
2.6.1 Modelo de Regressão Beta Burr XII	25
3 METODOLOGIA	27
3.1 Material	27
3.2 Métodos	27
4 RESULTADOS E DISCUSSÃO	29
4.1 Análise Exploratória dos dados de Suplementação de Vitamina A	29
4.2 Análise das Distribuições Aplicadas aos dados de Suplementação de Vitamina A	32
4.3 Ajuste dos Dados ao Modelo de Regressão Beta-Burr XII com Fração de Cura	34
4.4 Avaliação do Ajuste aos Dados dos Modelos de Regressão Locação e Escala Encaixados	36
4.5 Análise das Covariáveis	37
5 CONSIDERAÇÕES FINAIS	39
REFERÊNCIAS	41

ANEXOS	43
------------------	----

RESUMO

Modelo de regressão log-linear com fração de cura

Neste trabalho foi considerado diferentes modelos para representar uma situação experimental de análise de sobrevivência. O modelo de regressão Beta-Burr XII com fração de cura foi proposto para modelar os dados. Observou-se, ainda, o comportamento dos modelos de regressão encaixados Burr XII e Logístico. Os parâmetros do modelos foram estimados pelo método de máxima verossimilhança com restrição nos parâmetros. A comparação dos modelos se deu através do teste da razão de verossimilhança. O conjunto de dados utilizado se refere à um experimento realizado para avaliar o impacto da suplementação com vitamina A na redução da taxa da morbidade infantil por diarreia.

Palavras-chave: Modelos de regressão; Fração de cura; Beta-Burr XII; Burr XII; Log-Logístico.

1 INTRODUÇÃO

Em estudos estatísticos é possível encontrar dados em que a variável resposta é o tempo que um indivíduo levou para sofrer um evento. Esse tempo, conhecido como tempo de falha, ou tempo de sobrevivência, é estudado por meio de técnicas estatísticas da área de análise de sobrevivência. O termo *sobrevivência* está voltado para pesquisas na área médica, porém as mesmas metodologias podem ser aplicadas em outros setores como o industrial e financeiro. Assim, a análise de sobrevivência pode ainda ser conhecida como análise de confiabilidade.

Os dados de sobrevivência possuem algumas características especiais que os impedem de sofrer uma análise estatística comum. O primeiro aspecto a ser avaliado é a assimetria das observações. Os tempos de sobrevivência, geralmente, possuem uma assimetria à direita, excluindo a possibilidade de assumir normalidade na variável resposta do estudo. Além disso, em estudos de confiabilidade nem sempre é possível obter a informação de tempo de falha de todos os indivíduos, tendo assim uma informação incompleta da variável. Nesse último caso os dados são considerados dados censurados.

Uma observação pode ser dita censurada quando não se sabe o momento exato da falha, mas sabe-se que ela ocorreu após determinado tempo. Essa é a chamada censura à direita. Há outros mecanismos de censura, como censura à esquerda e censura intervalar. Neste trabalho será considerado apenas o mecanismo de censura à direita.

Os tempos de falha podem, ainda, estar relacionados à covariáveis. Para estudar o efeito dessas covariáveis é necessário aplicar um modelo de regressão capaz de acomodar dados censurados. Os modelos de regressão utilizados neste trabalho, são modelos paramétricos que assumem que os dados seguem uma distribuição de probabilidade conhecida.

Outra característica que pode estar presente nos dados de sobrevivência é a proporção de indivíduos curados. Isto é, diferentemente de indivíduos censuradas que sofrerão a falha em algum momento não observado, os indivíduos curados são aqueles não suscetíveis à falha e por conseguinte nunca irão falhar.

Dessa forma, tem-se então como objetivo principal ajustar um modelo de regressão com fração de cura para estudar um conjunto de dados de sobrevivência.

Todas as análises serão feitas com o auxílio do *software* R.

2 REVISÃO DE LITERATURA

2.1 Notação e conceitos básicos

A análise de sobrevivência é uma técnica estatística voltada para o estudo da ocorrência de eventos ao longo do tempo. Esses eventos são determinados de acordo com o interesse do pesquisador. Como sugerido pelo nome é comum o uso dessa técnica em estudos onde se pretende analisar o evento de morte de um determinado indivíduo ou produto.

Porém, o escopo de aplicabilidade desse método pode ser muito maior. Pode-se ter como exemplos casos onde o evento de interesse é o momento em que determinado indivíduo muda de cidade ou tempo em que um ex-presidiário leva para ser preso novamente. No setor econômico, pode se observar o tempo em que o indivíduo leva para se tornar inadimplente, ou quanto dura a ascensão de uma ação dentro do mercado.

O que diferencia a Análise de Sobrevivência das demais técnicas é sua capacidade de tratar dados com censura e dependência das covariáveis com o tempo. Entende-se por dados censurados, aqueles indivíduos que durante todo o período de observação do estudo não sofreu o evento de interesse. Se um estudo pretende avaliar a vida útil de uma determinada marca de lâmpadas, toma-se uma amostra e observa-se o comportamento dessas durante um período pré-definido. Ao fim desse período, provavelmente, nem todas as lâmpadas estarão queimadas e as lâmpadas que ainda funcionarem entrarão na base de dados como observações censuradas. Ou seja, não se sabe quanto tempo levou para elas queimarem, mas tem-se a informação de que esse tempo foi maior que o tempo de observação dos objetos. Nos métodos estatísticos convencionais essas informações são descartadas. Porém, há casos onde o número de dados censurados é muito grande, causando viés nos resultados caso eles sejam desconsiderados.

No estudo de análise de sobrevivência os dados consistem em tempos de eventos, observações censuradas e, em geral, de variáveis regressoras.

2.1.1 Tempo de falha

O tempo de falha é definido como tempo até a ocorrência do evento de interesse. Além disso, é constituído basicamente por três elementos: o tempo inicial, o evento de interesse (falha) e a escala de medida. Para que não haja viés nos resultados é importante

que estes três elementos estejam bem definidos.

O tempo inicial, já sugerido pelo nome, é caracterizado pelo instante em que se começa a contagem do tempo de observação do indivíduo. Geralmente esse tempo coincide com a entrada do indivíduo na amostra, entretanto, esse fato não é uma regra. O pesquisador pode ter em sua amostra indivíduos que obtiveram a cura do câncer e seu evento de interesse é a reincidência da doença. Nesse caso, seu tempo inicial será o momento em que o paciente obteve a cura e não necessariamente quando iniciou sua participação no estudo. Ao coletar informações sobre eventos já registrados, deve-se ter maior cautela para se definir o tempo inicial. Uma vez que se trata de eventos que podem ter ocorrido há longo tempo, um mau registro desse pode prejudicar a precisão das informações coletadas.

O evento de interesse é o objeto de estudo do pesquisador. Também conhecido como falha, o evento de interesse é o que se espera que aconteça com os indivíduos que estão sendo observados. Para que os resultados da pesquisa sejam válidos é importante que a falha esteja claramente definido. No caso da reincidência de câncer, por exemplo, é importante definir quais são as condições médicas que definem a presença de câncer no paciente. Uma vez identificadas todas as características que compõem o momento da falha é crucial que o mesmo critério seja adotado para todos os elementos do estudo.

A escala de medida se resume em definir se o tempo será medido em horas, dias, semanas, meses, anos, etc.. Assim como a definição do tempo inicial e do evento de interesse, é importante que a mesma escala de tempo seja usada para todos os indivíduos.

2.1.2 Censura

Um dos problemas encontrados no registro do evento de interesse é o fato desse nem sempre ocorrer durante o período de observação. Além disso, os elementos observados podem eventualmente deixar de fazer parte da pesquisa por algum motivo aleatório que não esteja associado ao evento de interesse. Nesses casos não será registrado o momento em que ocorreu a falha do indivíduo, mas poderá ser feito o registro do tempo mínimo em que o objeto de estudo permaneceu sem falhar. Essa informação, denominada censura, apesar de incompleta, é útil e importante para a análise.

As censuras podem ocorrer por diversos motivos. Há casos em que o pesquisador

determina o número de falhas necessárias para realizar suas análises. Uma vez atingido esse valor, os demais elementos da amostra deixam de ser observados e tornam-se dados censurados. É comum que isso aconteça em pesquisas industriais já que não é de interesse da indústria deixar todos os seus produtos falharem. Outro motivo que pode gerar observações censuradas é o fim da pesquisa, que geralmente é pré-definido, antes que ocorra a falha de todos os indivíduos observados. Além disso durante o estudo o pesquisador pode deixar de acompanhar alguns elementos da amostra que deixam de fazer parte da pesquisa por algum motivo extrínseco. Ainda, há os casos em que os indivíduos sofrem a falha, porém por motivos que não se relacionam com o fato que está sendo observado. Quando se estuda o tempo de morte de pacientes com determinada enfermidade e algum desses morre por acidente automobilístico, esse dados serão considerados uma censura.

A partir dos motivos que podem gerar um dado censurado, há na literatura a citação de três diferentes tipos de censura. Colosimo e Giolo (2006), indicam que o primeiro tipo de censura, ou seja, Censura do Tipo I, é quando o período de observações dos indivíduos se encerra é registrado que nem todos os elementos sofreram a falha. Neste caso os indivíduos sem registro de falha serão dados censurados e o momento que caracteriza o fim do estudo é pré-definido pelo pesquisador. A Censura do Tipo II é referida como aquela provocada pelo fim do estudo após ocorridos um determinado número de falhas. O outro tipo de censura, Censura do Tipo III, é conhecido como censura aleatória. Nesse último, é enquadrado os demais casos, onde não há a manipulação do pesquisador e por variados motivos a censura ocorre, de forma natural, nos dados. Essas três situações são referidas como mecanismo de censuras à direita, isto é, o momento em que ocorre o evento de interesse está à direita do tempo registrado.

Há ainda o mecanismo de censura à esquerda e censura intervalar. A primeira acontece quando o evento de interesse já ocorreu na vez em que o indivíduo é abordado, ou seja, o tempo de falha está à esquerda do tempo registrado. A censura intervalar é quando os indivíduos não podem ser acompanhados continuamente e obtêm-se apenas a informação de que a falha ocorreu em um determinado intervalo de tempo.

2.1.3 Representado o tempo de sobrevivência

O tempo de sobrevivência, caracterizado como uma variável aleatória, pode ter seu comportamento definido por diferentes funções. Algumas delas são: função densidade de probabilidade, função de sobrevivência e a função de risco. Todas estas funções possuem uma relação entre si, ou seja, ao identificar uma delas as demais podem ser derivadas.

A função densidade de probabilidade (f.d.p) é basicamente definida como o limite da probabilidade da falha ocorrer em um intervalo de tempo $[t, t+\Delta t)$, onde $\Delta t \rightarrow 0$. Pode-se expressar a função como:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

em que $f(t) \geq 0$ para todo t a área abaixo da curva igual a 1.

Além disso, essa função pode também ser derivada da função de distribuição acumulada, definida como a probabilidade da variável tempo de sobrevivência ser menor ou igual ao um determinado tempo, ou seja, $F(t) = P(T \leq t)$. Tem-se, então:

$$f(t) = \frac{dF(t)}{dt}. \quad (1)$$

A função de sobrevivência, $S(t)$, é o complementar da função de distribuição acumulada. Ela indica a probabilidade do indivíduo não falhar até um tempo t definido e é exposta por:

$$S(t) = P(T \geq t)$$

e

$$S(t) = 1 - P(T \leq t) = 1 - F(t) \quad (2)$$

A função de risco, também conhecida como taxa de falha, representa o risco instantâneo no tempo t , dado que o indivíduo sobreviveu até este tempo. Em relação às demais funções, a função de risco apresenta vantagens como a possibilidade de descrever a forma com que a taxa instantânea de falha muda com o decorrer do tempo. Além disso, ela é útil para representar a distribuição do tempo de vida. A taxa de falha é definida como a razão do limite da probabilidade de um indivíduo falhar no intervalo de tempo $[t, t+\Delta t)$,

dado que este não apresentou o evento de interesse até o tempo t sobre Δt . Desta forma, a função é expressa por:

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

A função de risco pode ser constante, crescente, decrescente ou mesmo não monótona. A partir dessas informações pode se atribuir ao tempo de sobrevivência diferentes distribuições como a Exponencial, Weibull, Log-Normal, entre outras.

Como já mencionado, todas as funções apresentadas possuem relações matemáticas entre si. Algumas das mais importantes funções definidas em Colosimo e Giolo (2006), são:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \left(\log S(t) \right)$$

e

$$S(t) = \exp \left\{ - \int_0^t h(u) du \right\}.$$

Cabe ressaltar que todas essas relações são válidas, apenas para quando a variável *tempo* é contínua.

2.1.4 Estimador de Kaplan Meier

O estimador de Kaplan Meier é uma técnica utilizada para obter as primeiras informações fornecidas pelo banco de dados. A presença de censuras impede que o estudo inicial seja feito a partir de uma análise descritiva comum. O estimador desenvolvido por Kaplan e Meier (1958), é uma técnica não-paramétrica que considera a presença de dados censurados e é uma solução para fazer uma espécie de análise exploratória dos dados. A ideia principal é estimar a função de sobrevivência e a partir dela estimar as estatísticas de interesse tais como tempo médio, mediano, frações de falhas em intervalos fixos de tempos, entre outras.

Ao definir o estimador de Kaplan Meier, Colosimo e Giolo (2006), afirmam a respeito desse: "Ele é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\widehat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não falharam até o tempo } t}{\text{n}^\circ \text{ total de observações no estudo}}.$$

O estimador não-paramétrico de Kaplan Meier é construído em cima de probabilidades condicionais. Por exemplo, para que um indivíduo sobreviva à 3 semanas, ele necessariamente precisa sobreviver à 1 e 2 semanas. Essa relação indica a condição de sobrevivência absorvida pelo estimador, que é expresso por:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right),$$

em que:

- $t_1 < t_2 < \dots < t_k$, são os k tempos de falhas, distintos e ordenados.
- d_j é o número de falhas em t_j , $j = 1, \dots, k$.
- n_j é a quantidade de indivíduos sob risco em t_j .

Segundo Colosimo e Giolo (2006), as principais propriedades do estimador de Kaplan Meier são:

- i) é não viciado para amostras grandes,
- ii) é fracamente consistente,
- iii) converge assintoticamente para distribuição normal e
- iv) é estimador de máxima verossimilhança de $S(t)$.

2.1.5 Função Total do Teste

Junto com o estimador de Kaplan-Meier pode-se utilizar outras técnicas para se ter o primeiro contato com os dados. A curva do tempo total em teste (TTT) foi proposta por Aarset(1987) e pode ser aplicada como uma técnica gráfica para identificar o modelo mais apropriado para os dados. O gráfico de TTT é obtido através de $G(r/n) = [(\sum_{i=1}^n T_{i:n} + (n - r)T_{r:n})/(\sum_{i=1}^n T_i : n)]$ por r/n , sendo $r = 1, \dots, n$ e $T_{i:n} i = 1, \dots, n$ as estatísticas de ordem da amostra.

Por meio desse gráfico é possível verificar o comportamento da função risco. A função risco pode ser constante, monótona, unimodal ou possuir a forma de banheira.

2.2 Distribuições de Probabilidade para os Dados

Neste trabalho serão considerados três diferentes tipos de modelo para acomodar os tempos de sobrevivência. São esses: Beta Burr XII, Burr XII e Log-logístico. A vantagem do uso dessas distribuições está ao testar a qualidade do ajuste, uma vez que as duas últimas são casos particulares da primeira.

Mais detalhes dessas distribuições aplicados à análise de sobrevivência serão descritos a seguir.

2.2.1 Distribuição Beta Burr XII (BBXII)

Como sugerido pelo nome a distribuição beta Burr XII é uma junção da distribuição beta com a distribuição Burr XII. A distribuição BBXII, acomoda diferentes tipos de função risco. Além das formas crescente, decrescente e unimodal, também acomodadas pela distribuição Burr XII, a BBXII, possui a forma de banheira ou U para a função risco.

Devido a sua flexibilidade para acomodar diferentes funções de risco, essa distribuição pode ser ajustada em variados problemas, tornando-se importante na análise de dados de sobrevivência.

Segundo Paranaíba (2012), a função de distribuição acumulada e a função densidade de probabilidade da BBXII, são respectivamente dadas por:

$$F(t) = \frac{1}{B(a, b)} \int_0^{1 - [1 + (\frac{t}{s})^c]^{-k}} \omega^{a-1} (1 - \omega)^{b-1} d\omega, t > 0$$

e

$$f(t) = \frac{ckt^{c-1}}{s^c B(a, b)} \left[1 + \left(\frac{t}{s} \right)^c \right]^{-(kb+1)} \left\{ 1 - \left[1 + \left(\frac{t}{s} \right)^c \right]^{-k} \right\}^{a-1}, t > 0. \quad (3)$$

A notação utilizada será $T \sim BBXII(a, b, s, k, c)$, em que $s > 0$ é o parâmetro de escala e $a > 0$, $b > 0$, $k > 0$ e $c > 0$ são os parâmetros de forma.

Para $a = b = 1$, tem-se a distribuição Burr XII. Ainda, para $a = b = 1$, $s = m^{-1}$ e $k = 1$ a distribuição BBXII é reduzida à distribuição log-logística.

Paranaíba (2012), ainda define a função de risco da distribuição BBXXII, como:

$$h(t) = \frac{ckt^{c-1} [1 + (\frac{t}{s})^c]^{-(kb+1)} \{1 - [1 + (\frac{t}{s})^c]^{-k}\}^{a-1}}{s^c B(a, b) [1 - I_{\{1 - [1 + (\frac{t}{s})^c]^{-k}\}}(a, b)]}, t > 0,$$

em que:

$$I_{\{1-[1+(\frac{t}{s})^c]^{-k}\}}(a, b) = F(t) = \frac{1}{B(a, b)} \int_0^{1-[1+(\frac{t}{s})^c]^{-k}} \omega^{a-1}(1-\omega)^{b-1} d\omega, t > 0$$

2.2.2 Distribuição Burr XII (BXIII)

A distribuição Burr XII é uma distribuição comum na modelagem de dados de sobrevivência. Com propriedades para acomodar funções de risco monótonas e unimodais a BXII, segundo Zimmer; Keats e Wang (1998), tem como função densidade de probabilidade:

$$f(t) = cks^{-c} \left[1 + \left(\frac{t}{s} \right)^c \right]^{(-k-1)} t^{c-1}, \quad (4)$$

em que os parâmetros de locação são $k > 0$ e $c > 0$ e o parâmetro de escala é representado por $s > 0$. Como notação será utilizado: $T \sim BXII(s, k, c)$.

A função de sobrevivência e função risco equivalente para o tempo de sobrevivência T com densidade Burr XII é definida, respectivamente, como:

$$S(t) = \left[1 + \left(\frac{t}{s} \right)^c \right]^{-k}$$

e

$$h(t) = \frac{ck(\frac{t}{s})^{c-1}}{s[1 + (\frac{t}{s})^c]}.$$

Algumas análises gráficas mostram que para valores de $c \leq 1$, $h(t)$ é decrescente, caso contrário a função de taxa de falha é unimodal. Ainda, pode-se verificar que para valores de c entre 1 e 2, $h(t)$ pode ser praticamente constante, dependendo do valor de s .

2.2.3 Distribuição Log-logística

A distribuição Log-logística é um caso particular das outras duas distribuições já apresentadas nessa seção. Colosimo e Giolo (2006), descreve sua função de densidade como:

$$f(t) = \frac{c}{s^c} t^{c-1} \left(1 + \left(\frac{t}{s} \right)^c \right)^{-2}, t > 0, \quad (5)$$

onde $s > 0$ é o parâmetro de escala e $c > 0$, parâmetro de locação.

As funções de sobrevivência e risco, são respectivamente expressas por:

$$S(t) = \frac{1}{1 + \left(\frac{t}{s}\right)^c}$$

e

$$h(t) = \frac{c\left(\frac{t}{s}\right)^{c-1}}{s[1 + \left(\frac{t}{s}\right)^c]}.$$

A função de taxa de falha, para valores de $c > 1$, tem comportamento crescente, atingindo um pico e em seguida decrescendo.

2.3 Estimação dos Parâmetros pelo Método de Máxima Verossimilhança

Em estudos de sobrevivência o interesse geral ao se estudar o comportamento de amostras é responder à perguntas de interesses envolvendo o caso analisado. O ajuste de um modelo a partir de dados amostrais permite ao pesquisador responder à perguntas que vão além do que foi observado no estudo.

Um modelo probabilístico, segundo Colosimo e Giolo (2006), é caracterizado por quantidades desconhecidas, os parâmetros. Os dados observados para uma amostra servem para que estes parâmetros possam ser estimados. No universo estatístico há variados métodos de estimação de parâmetros de modelos. Porém, nem todos são convenientes para estudos de tempos de vida, uma vez que não conseguem agregar as censuras nos seus procedimentos.

Um método de estimação capaz de incorporar as censuras é o Método de Máxima Verossimilhança. Sendo θ um único parâmetro ou um vetor destes, o método consiste, basicamente, em encontrar um valor para os parâmetros de tal forma que maximize a probabilidade da amostra observada ocorrer. Em outras palavras, Colosimo e Giolo (2006) afirma que o método procura encontrar uma distribuição que melhor explique a amostra observada.

Antes de obter a função de máxima verossimilhança para θ é importante analisar a natureza do mecanismo de censura, segundo Kalbfleisch e Prentice (2002). A função de máxima verossimilhança para tempos de vida é dividida em duas partes. Uma dessas é a função de densidade que representa a contribuição das observações não-censuradas. Já para os dados censurados suas contribuições são representadas pela sua função de sobrevivência $S(t)$. A menos de uma constante, Colosimo e Giolo (2006) definem a função de verossimilhança para todos os mecanismos de censuras à direita como:

$$L(\theta) \propto \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i},$$

em que:

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é um tempo de falha;} \\ 0 & \text{se } t_i \text{ é um tempo censurado.} \end{cases}$$

Os cálculos podem ser facilitados utilizando o logaritmo da função. Os estimadores obtidos serão os mesmos em qualquer das situações, uma vez que os valores encontrados serão aqueles que maximizam a função ou suas equivalentes. Assim, pode-se encontrar os estimadores que caracterizam o modelo resolvendo o seguinte sistema de equações:

$$U(\theta) = \frac{\partial \log(L(\theta))}{\partial \theta} = 0.$$

2.4 Teste da Razão de Verossimilhança

Para testar a qualidade do ajuste dos modelos supracitados pode ser realizado o teste de hipótese denominado por Teste da Razão de Verossimilhança. Sendo $\mathbf{x} = (x_0, \dots, x_p)'$ o vetor de parâmetros, é possível testar hipóteses relacionadas a este vetor ou a um subconjunto dele, segundo Colosimo e Giolo (2006).

A função de verossimilhança é a base deste teste. É feita a comparação dos valores dos logaritmos da função de verossimilhança maximizada sem restrição, no caso deste trabalho da Beta-Burr XII e sob H_0 , que seriam as log-verossimilhanças associadas às distribuições Burr XII e Log-Logística.

A estatística do teste é dada por:

$$TRV = -2 \log \left[\frac{L(\boldsymbol{\theta}_0)}{L(\boldsymbol{\theta})} \right] = 2 [\log(L(\boldsymbol{\theta})) - \log(L(\boldsymbol{\theta}_0))],$$

em que sob $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade, em que p corresponde à diferença entre o número de parâmetros do modelo completo e o número de parâmetros do modelo restrito. A hipótese nula será rejeitada, a um nível de significância de $100(1 - \alpha)\%$ quando $TRV > \chi_{p, 1-\alpha}^2$.

Pode-se ainda encontrar na literatura outros testes para verificação de adequabilidade. Os mais comuns são os Teste de Wald e o Teste Escore que são descritos brevemente por Colosimo e Giolo (2006).

2.5 Fração de Cura

Os modelos ajustados pela análise de sobrevivência, em geral, assumem que todos os elementos sob observação são suscetíveis à falha. Isto é, em um espaço de tempo suficientemente longo, os indivíduos sofrerão a falha. Nessas situações, caso a falha ocorra é registrado o tempo de falha, caso contrário a observação se torna censurada.

Na prática, a suposição de que todos os indivíduos são propensos à falha nem sempre será atendida. Segundo Fachini(2011), há situações em que uma parte da amostra pode não ser suscetível ao evento de interesse e dessa forma nunca experimentar o evento de falha. Essa proporção da amostra é conhecida como indivíduos *não suscetíveis* ou *curados*.

Os modelos de cura são modelos de sobrevivência que permitem alocar a fração de indivíduos curados. O primeiro passo para verificar se há presença de curados no dados é analisar a função de sobrevivência desses. Se for observado que essa função não converge para zero quando o tempo aumenta pode-se considerar a hipótese de presença de elementos não suscetíveis ao evento de interesse. Com esse tipo de comportamento a função de sobrevivência será conhecida como *imprópria*.

Para acomodar a fração de cura no modelo a população é dividida em duas subpopulações. Assim, ou o indivíduo é curado com probabilidade de $1 - \phi$ ou é propenso à falha ou censura com probabilidade ϕ , pertencendo ao grupo que compõe a função de sobrevivência $S(t)$. Um modelo para distribuição de tempos de sobrevivência que incorpora a fração de curados, segundo Berkson e Gage (1952), é dado por:

$$S^*(t) = (1 - \phi) + \phi S(t), \quad (6)$$

em que $\phi \in [0,1]$, com as seguintes propriedades: $\lim_{t \rightarrow \infty} S^*(t) = (1 - \phi)$ e $\lim_{t \rightarrow 0} S^*(t) = 1$.

Note que para os casos em que $\phi = 1$, tem-se como caso particular da equação (6) a função de sobrevivência própria.

Aplicando os resultados das equações (1) e (2) na equação (6), tem-se como resultado para a função densidade de probabilidade do modelo com fração cura:

$$\begin{aligned}
f^*(t) &= \frac{\partial F^*(t)}{\partial t} \\
&= \frac{\partial[1 - (1 - \phi) - \phi S(t)]}{\partial t} \\
&= \frac{\partial[-\phi S(t)]}{\partial t} \\
&= -\phi[-f(t)] \\
&= \phi f(t).
\end{aligned}$$

2.6 Modelo de Regressão Locação e Escala

No estudo de sobrevivência, é esperado que covariáveis influenciem o comportamento dos tempos de falha. Ainda, essas covariáveis podem explicar parte da heterogeneidade desses tempos. Em estudos na área médica, por exemplo, o tempo de sobrevivência de um paciente pode ser influenciado pela idade ou sexo do paciente.

Um meio para incluir as informações das covariáveis no estudo de dados de sobrevivência é o modelo de regressão. As covariáveis podem ser incluídas no modelo de diferentes formas. Neste trabalho será considerado apenas o modelo de locação e escala, em que as covariáveis são inseridas no parâmetro de locação.

Seja $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$ um vetor formado por observações de p variáveis regressoras. Considere, também, $Y = \log(T)$, ou seja, Y , como logaritmo dos tempos de sobrevivência. O modelo de locação e escala que relaciona Y e o vetor de variáveis explicativas \mathbf{x} é dado por:

$$Y = \mu(\mathbf{x}) + \sigma Z,$$

em que μ ($-\infty < \mu < \infty$) é o parâmetro de de locação, σ ($0 < \sigma < \infty$) é o parâmetro de escala e Z um erro aleatório com distribuição conhecida independente de \mathbf{x} .

A função densidade de probabilidade é definida por:

$$f(y; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{y - \mu(x)}{\sigma}\right), -\infty < y < \infty \quad (7)$$

e função de sobrevivência $G\left(\frac{y - \mu}{\sigma}\right)$.

Uma maneira comum de assumir $\mu(\mathbf{x})$ é:

$$\mu(\mathbf{x}) = \mathbf{x}^T \beta$$

em que $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ é o vetor de parâmetros desconhecidos associado às covariáveis.

Vale notar que o modelo é log-linear para os tempos de sobrevivência, T , e por isso um modelo de regressão linear para Y . Ainda, nota-se que o vetor de variáveis explicativas tem efeito multiplicativo em T .

2.6.1 Modelo de Regressão Beta Burr XII

Aplicando o resultado exposto acima, as funções densidade de probabilidade e de sobrevivência para $Y = \log(T)$ em que $T \sim BBXII(a, b, s, k, c)$, podem ser expressas, respectivamente, como:

$$f_Y(y) = \frac{k}{\sigma \text{Beta}(a, b)} \exp\left(\frac{y - \mu(x)}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-(kb+1)} \left\{1 - \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-k}\right\}^{a-1}$$

e

$$S(y) = 1 - \frac{1}{B(a, b)} \int_0^{1 - \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-k}} \omega^{a-1} (1 - \omega)^{b-1} d\omega,$$

em que $a, b, \sigma, k > 0$, $-\infty < y < \infty$ e $-\infty < \mu < \infty$.

Como já citado, as distribuições Burr XII e Log-logística são casos particulares da distribuição Beta Burr XII. Dessa forma, para $a = b = 1$, tem-se como função densidade de probabilidade e função de sobrevivência para o modelo de regressão Burr XII em Y :

$$f_Y(y) = \frac{k}{\sigma} \exp\left(\frac{y - \mu(x)}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-(k+1)}$$

e

$$S(y) = \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-k}.$$

em que $\sigma, k > 0$, $-\infty < y < \infty$ e $-\infty < \mu < \infty$.

A partir do modelo de regressão Beta Burr XII, assumindo $k = a = b = 1$, $\sigma > 0$, $-\infty < y < \infty$ e $-\infty < \mu < \infty$, obtém-se o modelo de regressão Log-Logístico que tem as seguintes funções de densidade de probabilidade e sobrevivência:

$$f_Y(y) = \frac{1}{\sigma} \exp\left(\frac{y - \mu(x)}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-2}$$

e

$$S(y) = \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right)\right]^{-1}.$$

Em todas as funções as covariáveis podem ser incluídas por meio da relação $\mu(x) = \mathbf{x}^T \beta$, em que $\beta = (\beta_0, \dots, \beta_p)$ é o vetor de parâmetros.

O desenvolvimento dos cálculos das funções de densidade de probabilidade para os modelos de regressão locação e escala, descritos nesta seção, podem ser encontrados no Anexo A. As funções de sobrevivência são obtidas automaticamente a partir das equações (1) e (2).

3 METODOLOGIA

3.1 Material

Para aplicar o modelo proposto neste trabalho, será utilizado um conjunto de dados reais. Os dados são oriundos de um experimento realizado para avaliar o impacto da suplementação com vitamina A na redução da taxa da morbidade infantil por diarreia.

O desenho da pesquisa de suplementação de vitamina A em crianças pré-escolares se fez em cima de um ensaio clínico de base comunitária, aleatório e placebo controlado. A coleta de dados durou 12 meses (dezembro de 1990 a dezembro de 1991) e envolveu uma amostra de 1.240 crianças do nordeste brasileiro.

O evento de interesse é a ocorrência de um episódio de diarreia na criança. Dessa forma, o tempo se falha ficou definido como o tempo entre a primeira dose de vitamina A até a ocorrência do evento de interesse. A fim de explicar parte da heterogeneidade dos tempos de falha há ainda informações de três covariáveis: idade, tipo de tratamento adotado (placebo ou vitamina A) e sexo da criança. Para mais detalhes acerca dos dados consulte Barreto et al. (1994).

3.2 Métodos

Considerando as seções 2.5 e 2.6, propõe-se o modelo de Regressão locação e escala com fração de cura, para ajuste dos dados observados.

Através das equações (6) e (7), pode-se obter o seguinte resultado geral para função de sobrevivência e função densidade de probabilidade para o modelo de regressão locação e escala com fração de cura, em que $Y = \log(T)$:

$$S^*(y) = (1 - \phi) + \phi S(y) \quad (8)$$

e

$$f^*(y) = \phi f(y). \quad (9)$$

Em particular, para o modelo de regressão Beta Burr XII com fração de cura as funções de densidade de probabilidade e sobrevivência serão determinadas por:

$$f_Y(y) = \phi \left(\frac{k}{\sigma \text{Beta}(a, b)} \exp\left(\frac{y - \mu(x)}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right) \right]^{-(kb+1)} \left\{ 1 - \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right) \right]^{-k} \right\}^{a-1} \right)$$

e

$$S(y) = (1 - \phi) + \phi \left(1 - \frac{1}{B(a, b)} \int_0^{1 - \left[1 + \exp\left(\frac{y - \mu(x)}{\sigma}\right) \right]^{-k}} \omega^{a-1} (1 - \omega)^{b-1} d\omega \right),$$

em que $a, b, \sigma, k > 0$, $-\infty < y < \infty$, $-\infty < \mu < \infty$ e $0 < \phi < 1$.

Para obter os casos particulares do modelo de regressão Beta-Burr XII basta tomar, $a = b = 1$ e será encontrado o modelo de regressão Burr XII com fração de cura. Para $k = a = b = 1$, tem-se o modelo de regressão logístico com fração de cura.

A estimação dos parâmetros será feita por meio do método de máxima verossimilhança restrita. Sendo o vetor de parâmetros θ sob w restrições de inequações lineares $\mathbf{u}_j^T \theta - c_j \geq 0$, $j = 1, 2, \dots, w$, em que \mathbf{u}_j^T é um vetor $w \times 1$ e c_j são escalares assumindo valores 0 ou 1, variando de acordo com a restrição de interesse. O vetor \mathbf{u}_j é do tipo $(1, 0, 0, \dots, 0)$ de tamanho w , assumindo valor 1 na posição em que o parâmetro de interesse se encontra.

A representação do logaritmo da função de verossimilhança sujeito às restrições lineares é dada por:

$$l_R(\theta, \vartheta) = l(\theta) + \vartheta \sum_{j=1}^w (\mathbf{u}_j^T \theta - c_j), \quad (10)$$

em que $\vartheta > 0$ é o parâmetro de ajuste e $(\mathbf{u}_j^T \theta - c_j)$ é o conjunto de restrições de inequações lineares.

Esse método de estimação é utilizado considerando o método da função barreira adaptada. Para maiores detalhes consultar (LANGE, 1999).

Neste trabalho, as estimativas de máxima verossimilhança com restrição nos parâmetros serão feitas com o auxílio do *software* R (R Development Core Team, 2012), através do pacote *constrOptim*.

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória dos dados de Suplementação de Vitamina A

Para análise descritiva dos dados será feita a estimativa de Kaplan-Meier para a função de sobrevivência e a construção do gráfico da curva Tempo Total em Teste para avaliar o comportamento da função risco. Além disso, também serão feitas análises para avaliar o comportamento de cada grupo das covariáveis presentes no banco de dados.

Na Figura 1, através da representação gráfica da estimativa de Kaplan-Meier, nota-se que a probabilidade da criança não ter diarreia até 135 dias após receber a dosagem de vitamina A ou placebo, torna-se estável em aproximadamente 22%. Esse fato é um indicativo da presença de indivíduos imunes, o que justifica o uso de um modelo com fração de cura. Além disso, pode-se observar que maior parte dos dados censurados ocorrem a partir de 120 dias de estudo.

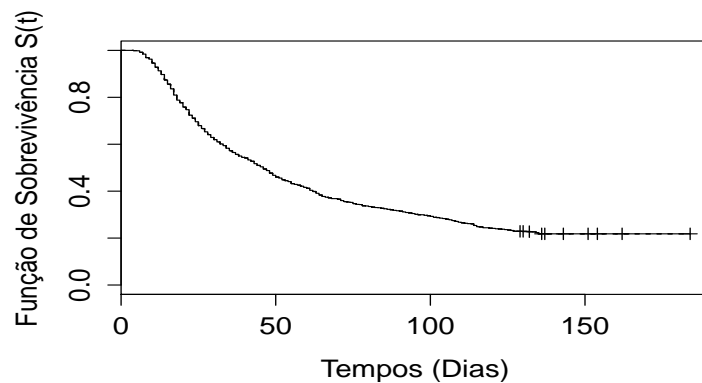


Figura 1 – Curva de sobrevivência estimada pelo método de Kaplan-Meier para os tempos de falha.

Para avaliar o impacto do sexo das crianças nos tempos de falha foram feitas as estimativas de Kaplan-Meier para cada grupo. Ambas as curvas não aparentam ser distintas. O resultado de cada gênero é ilustrado na Figura 2.

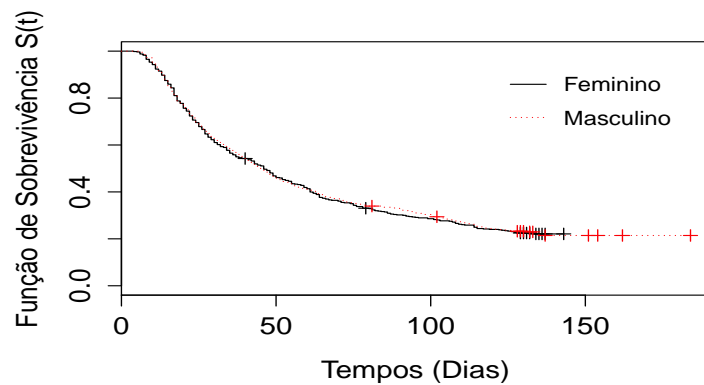


Figura 2 – Curva de sobrevivência estimada pelo método de Kaplan-Meier para os tempos de falha de cada gênero.

A covariável *tratamento* é analisada na Figura 3, em que nota-se uma sutil diferença entre as curvas de sobrevivência. Essa suposta diferença indica que as crianças que receberam dose de vitamina A tem maior probabilidade de sobrevivência que as demais crianças, dado um mesmo tempo t .

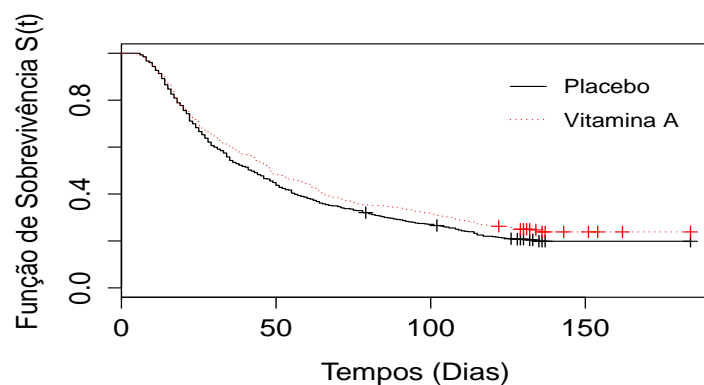


Figura 3 – Curva de sobrevivência estimada pelo método de Kaplan-Meier para os tempos de falha de cada tratamento.

Para determinar o comportamento da função risco, observa-se a Figura 4. Pode-se notar um comportamento crescente, tornando-se estável e em seguida crescendo novamente.

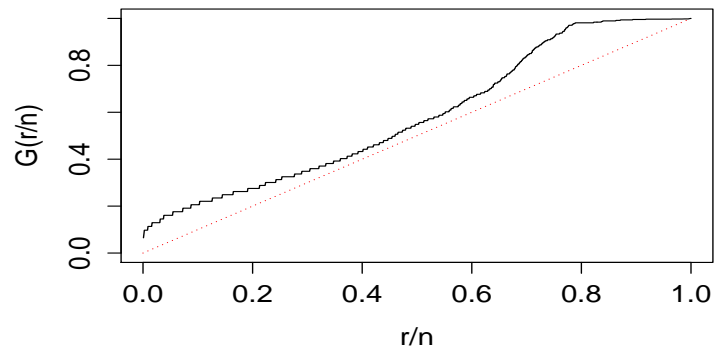


Figura 4 – Tempo total em teste para os tempos de falha.

Analisando o comportamento dos indivíduos de diferentes sexo e submetidos a diferentes tipo de tratamento, nota-se nas Figuras 5 e 6 a mesma tendência de distribuição para ambas as covariáveis e sutis diferenças entre sexo feminino, masculino e tratamento placebo, vitamina A.

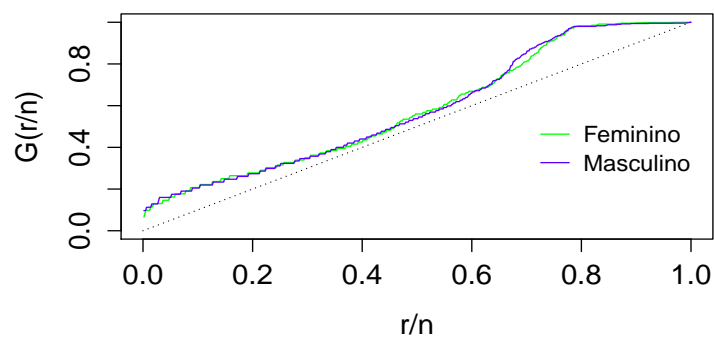


Figura 5 – Tempo total em teste para os tempos de falha por sexo.

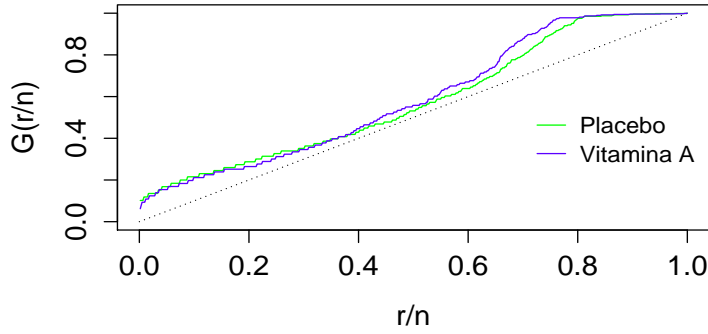


Figura 6 – Tempo total em teste para os tempos de falha por tipo de tratamento.

Esses resultados sugerem uma possível distribuição bimodal para os dados. Porém, pode-se esperar com a inclusão das informações das covariáveis, que uma distribuição unimodal seja adequada para modelagem dos dados. Assim, será desenvolvido um modelo de regressão com as distribuições descritas na Seção 2.2 e uma extensão desses, ao incluir a informação de possíveis indivíduos curados no estudo. Tudo isso, originará um modelo de regressão com fração de cura.

4.2 Análise das Distribuições Aplicadas aos dados de Suplementação de Vitamina A

De acordo com os resultados encontrados na seção anterior, serão avaliadas algumas distribuições que acomodam o comportamento unimodal dos tempos de falha. São essas distribuições: Beta Burr XII, Burr XII e Log-logística, já apresentadas neste trabalho.

Com o auxílio do *software* R, os parâmetros para as três distribuições foram estimados e podem ser verificados abaixo:

Tabela 1 – Estimativa dos parâmetros dos modelos e o valor de sua verossimilhança.

Modelo	Estimativas					Verossimilhança
	s	c	k	a	b	
Beta Burr XII	4,53	2,94	0,24	4,23	1,27	-4836,66
Burr XII	14,55	3,45	0,19	-	-	-4843,60
Log-logística	48,56	1,49	-	-	-	-4910,17

Para verificar qual modelo melhor se ajusta aos dados, foi realizado o teste da razão de verossimilhança e uma análise gráfica, comparando a curva dos modelos com a curva estimada pelo método de Kaplan-Meier.

O resultado do teste da razão de verossimilhança (TRV) pode ser verificado na Tabela 2. O teste mostra que estatisticamente o modelo mais completo, Beta Burr XII, ajusta melhor os dados que os demais modelos encaixados.

Tabela 2 – Teste da Razão de Verossimilhança.

Modelo	$\log(L(\boldsymbol{\theta}))$	TRV	g.l.	valor-p
Beta Burr XII	-4836,66	-	-	-
Burr XII	-4843,60	13,88	2	<0.0001
Log-logística	-4910,17	133,15	1	0

Fazendo uma análise gráfica a partir da Figura 7 confirma-se o resultado exibido na Tabela 2, uma vez que a curva de sobrevivência ajustada pela distribuição Beta Burr XII está notoriamente mais próxima da estimativa não paramétrica de Kaplan - Meier.

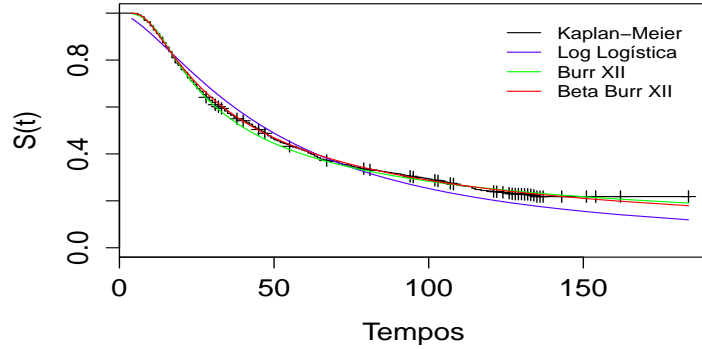


Figura 7 – Curvas de sobrevivência para os tempos de falha.

A partir do resultado evidenciado pela Tabela 2 e Figura 1, a análise do modelo de regressão locação e escala com fração de cura será feita a partir da distribuição Beta Burr XII.

4.3 Ajuste dos Dados ao Modelo de Regressão Beta-Burr XII com Fração de Cura

Para o ajuste do modelo de regressão Beta-Burr XII com fração de cura, acrescentou-se as informações das covariáveis: idade, tratamento e sexo, sendo seus coeficientes respectivamente: β_1 , β_2 , β_3 e tendo ainda o β_0 como intercepto. Além disso, tem-se o parâmetro ϕ que representa a proporção de indivíduos suscetíveis à falha, e por conseguinte, $(1 - \phi)$ representa a proporção de indivíduos curados.

Os resultados das estimativas dos modelos com e sem fração de cura estão exibidos na Tabela 3.

Tabela 3 – Estimativa dos parâmetros dos Modelos de Regressão Beta Burr XII com fração de cura e sem fração de cura e o valor de suas verossimilhanças.

Parâmetros do Modelo	Estimativas	
	Modelo Com Fração de Cura	Modelo Sem fração de cura
ϕ	0,99	-
σ	0,60	0,61
k	0,20	0,22
a	5,23	5,27
b	4,15	3,92
β_0	0,74	0,74
β_1	0,02	0,02
β_2	0,08	0,08
β_3	0,06	0,06
Verossimilhança	-1604,24	-1604,26

A Tabela 3 indica que o parâmetro que representa a proporção de indivíduos curados no estudo se aproxima do valor 1, isto é, há evidências de que o parâmetro ϕ é não-significativo. Para testar a significância do parâmetro na fronteira do espaço paramétrico, segundo Maller e Zhou (1996), deve-se assumir que $2(\log(L(\boldsymbol{\theta})) - \log(L(\boldsymbol{\theta}_0)))$ tem distribuição Qui-quadrada definida por: $P(X \leq x) = \frac{1}{2} + \frac{1}{2}P(\chi^2 \leq x)$. O percentil de 95% desta distribuição é 2,71. Caso o valor da estatística do teste seja menor que 2,71, não se rejeita a hipótese nula, de que $\phi = 1$. No caso dos modelos expostos na Tabela 3, tem-se: $2(-1604,24 + 1604,26) = 0,02 < 2,71$. Portanto, o resultado revela que não há evidências para rejeitar a hipótese nula de que ambos os modelos são equivalentes. Ainda, tem-se que as estimativas dos parâmetros tanto em um modelo, quanto no outro são bem próximas.

Com este resultado, descarta-se a hipótese de que neste estudo há indivíduos não suscetíveis à falha. Isto é, crianças imunes à diarreia.

4.4 Avaliação do Ajuste aos Dados dos Modelos de Regressão Locação e Escala Encaixados

Após detectar que não há evidências para afirmar a presença de indivíduos curados na amostra, é necessário reavaliar a significância dos demais parâmetros. Para isso, será inicialmente testada a hipótese dos parâmetros a e b serem iguais à 1, isto é, $a = b = 1$, tendo como resultado caso a hipótese não seja rejeitada, o modelo de regressão Burr XII. Para isso, é feito novamente o TRV comparando o Modelo de Regressão Beta Burr XII com o Modelo de Regressão Burr XII.

Os resultados das estimativas dos parâmetros e do teste da razão de verossimilhança são apresentados nas tabelas 4 e 5, respectivamente.

Tabela 4 – Estimativa dos parâmetros dos Modelos de Regressão Locação e Escala e o valor de suas verossimilhanças.

Parâmetros do Modelo	Estimativas	
	Modelo de Regressão Beta Burr XII	Modelo de Regressão Burr XII
a	5,27	1
b	3,92	1
σ	0,61	0,36
k	0,22	0,28
β_0	0,74	2,25
β_1	0,02	0,02
β_2	0,08	0,09
β_3	0,06	0,04
Verossimilhança	-1604,26	-1611,67

Tabela 5 – Teste da Razão de Verossimilhança para os Modelos de Regressão Locação e Escala.

Modelo	$\log(L(\boldsymbol{\theta}))$	TRV	g.l.	valor-p
Beta Burr XII	-1604,26	-	-	-
Burr XII	-1611,67	14,82	2	<0.0001

Novamente, o teste da razão de verossimilhança, indica que não há evidências de que o Modelo de Regressão Burr XII seja tão adequado para ajuste dos dados quanto o Modelo de Regressão Beta Burr XII. Portanto, assume-se o modelo mais completo como melhor modelo para explicar os tempos de falha e sua relação com as covariáveis.

4.5 Análise das Covariáveis

É importante verificar a significância das covariáveis presentes no modelo. Isto é, avaliar se elas influenciam no comportamento dos tempos de falha ou não.

A partir do modelo escolhido na seção anterior, Modelo de Regressão Beta - Burr XII, as covariáveis foram testadas retirando cada uma delas do modelo, uma a uma, e utilizando o logaritmo da função de verossimilhança para comparar os resultados, através do teste da razão de verossimilhança. A Tabela 6 apresenta os resultados dos testes comparando cada modelo sem uma das covariáveis, com o modelo completo.

O resultado do TRV exibido na Tabela 6 indica que o modelo de regressão Beta Burr XII sem a covariável *idade* não é estatisticamente equivalente ao modelo completo, indicando que esta covariável é significativa no modelo. O mesmo não é válido para as demais covariáveis, pois o TRV aponta que se as retirarmos, uma a uma, do modelo obtendo um modelo mais simples, este será tão adequado quanto o modelo completo.

Tabela 6 – Teste da Razão de Verossimilhança para testar covariáveis

Modelo	$\log(L(\boldsymbol{\theta}))$	TRV	g.l.	valor-p
Beta Burr XII	-1604,26	-	-	
Beta Burr XII (<i>sem covariável idade</i>)	-1637,53	66,54	1	$<0,0001$
Beta Burr XII (<i>sem covariável tratamento</i>)	-1605,25	1,98	1	0,16
Beta Burr XII (<i>sem covariável sexo</i>)	-1604,8	1,08	1	0,30

Para confirmar que o Modelo de Regressão Beta - Burr XII com a covariável *idade* é tão adequado quanto o modelo completo, fez-se novamente o teste da razão de verossimilhança para efeitos de comparação, conforme Tabela 7.

Tabela 7 – Teste da Razão de Verossimilhança para testar covariáveis

Modelo	$\log(L(\boldsymbol{\theta}))$	TRV	g.l.	valor-p
Beta Burr XII	-1604,26	-	-	
Beta Burr XII (<i>apenas com a covariável idade</i>)	-1605,89	3,26	2	0,20

Portanto não há evidências para rejeitar a hipótese de que o modelo mais simples apenas com a covariável idade é tão significativo para explicar o comportamento dos tempos de falha quanto o modelo completo. Assim, adotando o princípio da parcimônia adota-se como modelo o modelo mais simples. Ainda, de acordo com o modelo final tem-se que quanto mais velha for a criança, maior o tempo até que ela apresente sintomas de diarreia. Em relação ao gênero e tratamento recebido pela criança não houve evidências para assegurar a significâncias dessas covariáveis ao descrever o comportamento dos tempos de sobrevivência, como esperado a partir dos resultados observados nas Figuras 2 e 3.

5 CONSIDERAÇÕES FINAIS

A partir do estudo desenvolvido neste trabalho e dos resultados encontrados, conclui-se que o modelo com mais parâmetros, Modelo de Regressão Beta - Burr XII é o mais adequado para ajuste dos tempos de falha dos indivíduos observados. Porém, notou-se, também, que apenas a inclusão da covariável *idade* é suficiente para explicar o comportamento dos dados.

Como esperado, uma distribuição unimodal ajustou-se bem aos dados, tendo suas estimativas bem próximas às encontradas pelo método não-paramétrico de Kaplan-Meier. Porém, os resultados mostraram que não há evidências de indivíduos não suscetíveis à falha na amostra.

O modelo escolhido indicou que quanto mais velha for a criança maior será o tempo até que ela apresente sintomas de diarreia. Contudo, no estudo realizado não foi constatado que a vitamina A tem efeito significativo no tempo de falha da criança, isto é, não há diferença entre esta receber ou não doses da vitamina. O mesmo ocorreu quanto ao gênero dos indivíduos observados.

REFERÊNCIAS

- AARSET, M.V. How to identify bathtub hazard rate. **IEEE Transactions on Reliability**, New York, v.36, p.106-108,1987
- BARRETO, M.L.; SANTOS, L.M.P.; ASSIS, A.M.O.; ARAÚJO, M.P.N.; FARENZENA, G.G.; SANTOS, P.A.B.; FIACCONE, R.L. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. **Lancet**, London, v.344, p.228-231,1994
- BERKSON, J.; GAGE, R.P. Survival curve for cancer patients follwing treatment. **Journal of the American Statistical Association**, Alexandria, v.47, p. 501-515, 1952
- COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**, São Paulo: Blucher, 2006. 392p.
- FACHINI, J.B, **Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência**. 2011. 140p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.
- KALBFLEISCH, J.D.; PRENTICE, R.L. **The statistical analysis of failure time data**. 2nd ed. New York: John Wiley, 2002. 439p.
- KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. New York: Springer Verlag, 1997. 536 p.
- LANGE, K. **Numerical analysis for statisticians..** New York: Springer, 1999. 356p.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**, 2nd ed., New York: Wiley, 2003. 439 p.
- MALLER, R. A.; ZHOU, X. **Survival Analysis with Long-Term Survivors**. United Kingdom: John Wiley & Sons Ltd., 1996. 308p.
- PARANAÍBA, P.F. **Caracterização e extensões da distribuição Burr XII: propriedades e aplicações**. 2011. 142p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.
- R Core Team (2012). **R:A language and environment for statistical computing**, Vienna, Austria, Disponível em: <http://www.R-project.org/>
- SILVA, G.O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificada**. 2009. 207p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2009.
- ZIMMER, W.J.; KEATS, J.B.; WANG, F.K. The Burr XII distribution in reliability analysis. **Journal of Quality Technology**, Milwaukee, v.30, p. 389-394, 1998.

ANEXOS

Anexo A: Modelos de regressão locação e escala

A seguir, serão apresentados os cálculos dos modelos de regressão locação e escala para as distribuições Log-logística, Burr XII e Beta-Burr XII.

Modelo Log-logístico

Considere t uma variável aleatória com distribuição log-logística. Seja $Y = \log(t)$. Tem-se, então, que:

1. $t = \exp(y)$
2. $f_Y(y) = f_T(y)|J|$, em que:
 $|J| = \frac{dt}{dy} = \exp(y)$

Aplicando esses resultados juntamente com a Equação 5, tem-se que:

$$\begin{aligned} f_Y(y) &= \frac{c}{s^c} \exp(y)^{c-1} \left[1 + \left(\frac{\exp(y)}{s} \right)^c \right]^{-2} \exp(y) \\ &= \frac{c}{s^c} \exp(y)^c \left[1 + \left(\frac{\exp(y)^c}{s^c} \right) \right]^{-2} \\ &= \frac{c}{s^c} \exp(cy) \left[1 + \left(\frac{\exp(cy)}{s^c} \right) \right]^{-2}. \end{aligned}$$

Fazendo as seguintes substituições: $s = \exp(\mu)$ e $c = 1/\sigma$, temos:

$$\begin{aligned} f_Y(y) &= \frac{1/\sigma}{\exp(\mu)^{\frac{1}{\sigma}}} \exp(y/\sigma) \left[1 + \left(\frac{\exp(y/\sigma)}{\exp(\mu)^{1/\sigma}} \right) \right]^{-2} \\ &= \frac{1}{\sigma} \frac{\exp(y/\sigma)}{\exp(\mu/\sigma)} \left[1 + \left(\frac{\exp(y/\sigma)}{\exp(\mu/\sigma)} \right) \right]^{-2} \\ &= \frac{1}{\sigma} \exp\left(\frac{y - \mu}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-2}. \end{aligned}$$

Portanto, $Y \sim \text{Logística}(\mu, \sigma)$.

Modelo Burr XII

Seja t uma variável aleatória com distribuição Burr XII. Considere ainda que, $Y = \log(t)$. Seguindo os passos do caso anterior e utilizando a Equação 4, tem-se que:

$$\begin{aligned} f_Y(y) &= ck \left[1 + \left(\frac{\exp(y)}{s} \right)^c \right]^{(-k-1)} \left(\frac{\exp(y)^{c-1}}{s^c} \right) \exp(y) \\ &= ck \left[1 + \frac{\exp(cy)}{s^c} \right]^{(-k-1)} \frac{\exp(cy)}{s^c}. \end{aligned}$$

Fazendo as seguintes substituições: $s = \exp(\mu)$ e $c = 1/\sigma$, temos:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} k \left[1 + \frac{\exp(\frac{1}{\sigma} y)}{\exp(\mu)^{1/\sigma}} \right]^{-(k+1)} \frac{\exp(\frac{1}{\sigma} y)^{1/\sigma}}{\exp(\mu)} \\ &= \frac{k}{\sigma} \exp\left(\frac{y - \mu}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-(k+1)}. \end{aligned}$$

Portanto, $Y \sim \log - BurrXII(\mu, \sigma, k)$.

Modelo Beta Burr XII

Suponha, agora, que t é uma variável aleatória com distribuição Beta Burr XII.

Analogamente aos casos anteriores e fazendo uso da Equação 3, tem-se que:

$$f_Y(y) = \frac{c k \exp(y)^{c-1}}{s^c B(a, b)} \left[1 + \frac{\exp(y)^c}{s^c} \right]^{-(kb+1)} \left\{ 1 - \left[1 + \left(\frac{\exp(y)}{s} \right)^c \right]^{-k} \right\}^{a-1} \exp(y).$$

Fazendo as seguintes substituições: $s = \exp(\mu)$ e $c = 1/\sigma$, temos:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} \frac{k \exp(y)^{1/\sigma}}{\exp(\mu) B(a, b)} \left[1 + \frac{\exp(y)^{1/\sigma}}{\exp(\mu)^{1/\sigma}} \right]^{-(kb+1)} \left\{ 1 - \left[1 + \left(\frac{\exp(y)}{\exp(\mu)} \right)^{1/\sigma} \right]^{-k} \right\}^{a-1} \\ &= \frac{k}{\sigma B(a, b)} \exp\left(\frac{y - \mu}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-(kb+1)} \left\{ 1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-k} \right\}^{a-1}. \end{aligned}$$

Portanto, $Y \sim \log - BetaBurrXII(\mu, \sigma, k)$.